

CLAIMS

What is claimed is:

1. A computer implemented method of extracting information from a database,
5 comprising:
searching the database for occurrences of at least one tuple of information;
analyzing an occurrence of a tuple of information that was found in the database to
identify a pattern in which the tuple of information was stored; and
extracting additional tuples of information from the database utilizing the pattern.
10
2. The method of claim 1, further comprising providing the at least one tuple of
information as an example of information that is desired.
3. The method of claim 1, repeating the searching, analyzing and extracting for
15 additional tuples of information.
4. The method of claim 3, wherein the repeating the searching, analyzing and
extracting for the additional tuples of information continues until a predetermined number of
tuples of information are extracted.
20
5. The method of claim 1, wherein the pattern is defined by a regular expression,
context free grammar or computable function.
6. The method of claim 1, wherein the pattern includes a middle text, where the
25 middle text is between desired information in the tuples of information.
7. The method of claim 1, wherein the pattern includes a prefix text and suffix
text, where the prefix text precedes desired information in the tuples of information and the
suffix text follows desired information in the tuples of information.

8. The method of claim 1, wherein the pattern includes an order of the information in each tuple of information.

5 9. The method of claim 1, wherein the pattern includes a URL prefix, where the URL prefix is the initial portion of the URL where the pattern was identified.

10 10. The method of claim 1, further comprising verifying if an additional tuple of information matches a predetermined number of patterns, wherein the predetermined number of patterns is greater than 1.

11. The method of claim 10, wherein the additional tuple is rejected if it does not match at least the predetermined number of patterns.

15 12. The method of claim 1, further comprising verifying if the pattern has a specificity less than a predetermined specificity.

20 13. The method of claim 12, wherein the pattern is rejected if the specificity less than the predetermined specificity.

25 14. The method of claim 12, further comprising repeating the searching, analyzing and extracting for additional tuples of information, wherein the searching, analyzing and extracting for the additional tuples of information continues until no more patterns are identified that have a specificity greater than the predetermined specificity.

15. The method of claim 12, further comprising calculating the specificity by multiplying text string lengths of components of the pattern.

30 16. The method of claim 12, wherein the specificity increases in proportion to the number of tuples of information that match the pattern.

17. The method of claim 1, wherein the database is the World Wide Web.

18. A computer program product for extracting information from a database, comprising:

5 computer code that searches the database for occurrences of at least one tuple of information;

computer code that analyzes an occurrence of a tuple of information that was found in the database to identify a pattern in which the tuple of information was stored;

computer code that extracts additional tuples of information from the database
10 utilizing the pattern; and

a computer readable medium that stores the computer codes.

19. The computer program product of claim 18, wherein the computer readable medium is a CD-ROM, floppy disk, tape, flash memory, system memory, hard drive, or data
15 signal embodied in a carrier wave.

20. A computer implemented method of extracting information from a database, comprising:

searching the database for occurrences of tuples of information;

20 analyzing the occurrences of the tuples of information that were found in the database to identify a pattern in which the tuples of information were stored, wherein a pattern includes a prefix text, a middle text and suffix text, where the prefix text precedes desired information in the tuples of information, the middle text is between desired information in the tuples of information and the suffix text follows desired information in the tuples of information;

25 extracting additional tuples of information from the database utilizing the pattern; and repeating the searching, analyzing and extracting for additional tuples of information.

21. The method of claim 20, further comprising providing the tuples of information as examples of information that are desired.

30

22. The method of claim 20, wherein the repeating the searching, analyzing and extracting for the additional tuples of information continues until a predetermined number of tuples of information are extracted.

5 23. The method of claim 20, wherein the pattern includes an order of the information in the tuples of information.

24. The method of claim 20, wherein the pattern includes a URL prefix, where the URL prefix is the initial portion of the URL where the pattern was identified.
10

25. The method of claim 20, further comprising verifying if an additional tuple of information matches a predetermined number of patterns, wherein the predetermined number of patterns is greater than 1.

15 26. The method of claim 25, wherein the additional tuple is rejected if it does not match at least the predetermined number of patterns.

27. The method of claim 20, further comprising verifying if the pattern has a specificity less than a predetermined specificity.
20

28. The method of claim 27, wherein the pattern is rejected if the specificity less than the predetermined specificity.

29. The method of claim 27, further comprising repeating the searching, analyzing
25 and extracting for additional tuples of information, wherein the searching, analyzing and extracting for the additional tuples of information continues until no more patterns are identified that have a specificity greater than the predetermined specificity.

30 30. The method of claim 27, further comprising calculating the specificity by multiplying text string lengths of components of the pattern.

31. The method of claim 27, wherein the specificity increases in proportion to the number of tuples of information that match the pattern.

32. The method of claim 20, wherein the database is the World Wide Web.

5

33. A computer program product for extracting information from a database, comprising:

computer code that searches the database for occurrences of tuples of information;

computer code that analyzes the occurrence of tuples of information that were found in the database to identify a pattern in which the tuples of information were stored, wherein a pattern includes a prefix text, a middle text and suffix text, where the prefix text precedes desired information in the tuples of information, the middle text is between desired information in the tuples of information and the suffix text follows desired information in the tuples of information;

computer code that extracts additional tuples of information from the database utilizing the pattern;

computer code that repeats the searching, analyzing and extracting for additional tuples of information; and

a computer readable medium that stores the computer codes.

20

34. The computer program product of claim 33, wherein the computer readable medium is a CD-ROM, floppy disk, tape, flash memory, system memory, hard drive, or data signal embodied in a carrier wave.